

# RADIO ASTRONOMICAL IMAGE DECONVOLUTION USING PROLATE SPHEROIDAL WAVE FUNCTIONS

Sarod Yatawatta

Kapteyn Astronomical Institute, University of Groningen, Groningen,  
and ASTRON, Dwingeloo,  
The Netherlands.

## ABSTRACT

In order to produce high dynamic range images in radio interferometry, bright extended sources need to be removed with minimal error. However, this is not a trivial task because the Fourier plane is sampled only at a finite number of points. The ensuing deconvolution problem has been solved in many ways, mainly by algorithms based on CLEAN. However, such algorithms that use image pixels as basis functions have inherent limitations and by using an orthonormal basis that span the whole image, we can overcome them [1]. The construction of such an orthonormal basis involves fine tuning of many free parameters that define the basis functions. The optimal basis for a given problem (or a given extended source) is not guaranteed. In this paper, we discuss the use of generalized prolate spheroidal wave functions as a basis. Given the geometry (or the region of interest) of an extended source and the sampling points on the visibility plane, we can construct the optimal basis to model the source. Not only does this give us the minimum number of basis functions required but also the artifacts outside the region of interest are minimized.

**Index Terms**— Radio astronomy, Radio interferometry, Deconvolution

## I. INTRODUCTION

Due to the increase in computing power, high dynamic range imaging in radio interferometry (only limited by calibration errors) is achievable and is essential to produce novel scientific results. One of the main obstacles to reach a high dynamic range is the fact that only a finite number of samples of the Fourier (visibility) plane data is observed. Moreover, in earth rotation synthesis, these sampling points are not on a regular grid. If the observed field of view contains bright extended sources, complete removal (deconvolution) of such sources to reveal the faint background remains a challenge. The commonly used algorithm for such problems is CLEAN [2]. However, as shown in [1] (and references therein), the usage of a set of image pixels (clean components) in CLEAN based algorithms has limitations.

An alternative approach to using clean components is to use an orthonormal basis to model bright extended structure in the observation. Once such a model is obtained, it can be subtracted from the visibilities to reveal the residual (or the fainter background). The most obvious method of deriving such a basis is to use the observed data itself [3]. However, this has the drawback that the derivation of the basis can only be done once the complete observation is available (problematic for real time imaging etc.). On the other hand, we can adopt any arbitrary orthonormal basis to a given observation, completely independent of the data. For instance, in [1] Gauss-Hermite polynomials (shapelets) were used to produce high dynamic range images. The drawback in such an approach is the selection of free parameters (such as the number of basis functions, and the scale) cannot be determined in an optimal fashion. It requires experience of the user as well as a trial and error approach to fine tune such a basis for a given observation.

Due to the noise floor, any extended source has finite support [4], and the region of interest (ROI) or the support of a given extended source might not be optimal for a given arbitrary basis.

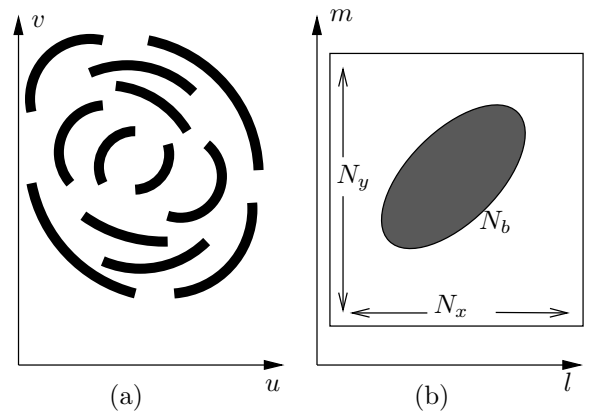
In order to tackle the problem of finding the optimal basis for a given extended source, independent of the observed data, we select prolate spheroidal wave functions (PSWF) [5], [6]. A similar problem has been solved in magnetic resonance imaging [7] and we extend that result to radio interferometry in this paper. Unlike data derived basis functions, PSWF basis can be precomputed and can be reused for observations at different epochs. Furthermore, unlike shapelets, we suffer less from artifacts outside the ROI with minimal number of basis functions used. In fact, PSWF are already being used in radio astronomical imaging to construct a regular grid of sampling points in the Fourier plane [8]. We refer the reader to [9] for similar applications of PSWF in geoscience.

Notation: We denote vectors in bold lowercase and matrices in bold uppercase. The matrix transpose, Hermitian, pseudoinverse are denoted by  $(\cdot)^T$ ,  $(\cdot)^H$  and  $(\cdot)^\dagger$  respectively. The identity matrix is given by  $\mathbf{I}$ .

## II. MATHEMATICAL FOUNDATIONS

We present the basics of interferometric imaging and the use of PSWF in this section. For a complete overview of radio interferometry, the reader is referred to [8].

### II-A. Interferometric imaging



**Fig. 1.** (a) Sampling points (total  $N_a$ ) in the Fourier (visibility) plane. (b) Image of  $N_x$  by  $N_y$  pixels, with the support (ROI) area shaded. The shaded area has  $N_b$  pixels.

We consider the visibility plane to be composed of  $N_a$  sampling points as in Fig. 1 (a). The image has  $N$  ( $= N_x \times N_y$ ) pixels. However, the ROI has only  $N_b$  pixels, corresponding to the shaded

area in Fig. 1 (b). The ROI is determined by the support of the source structure and the noise floor [4].

Let  $f(u_p, v_p)$  be the sampled visibility at the  $p$ -th point in the visibility plane. Let us also denote the intensity of the  $q$ -th pixel on the image as  $f(l_q, m_q)$ . These two quantities are related by the van Cittert-Zernike theorem [8] and can be approximated by the Fourier transform for images with small support as:

$$f(l_q, m_q) = \sum_{p=0}^{N_a-1} \tilde{f}(u_p, v_p) e^{j2\pi(l_q u_p + m_q v_p)}, \quad q \in [0, N-1] \quad (1)$$

$$\tilde{f}(u_p, v_p) = \sum_{q=0}^{N-1} f(l_q, m_q) e^{-j2\pi(l_q u_p + m_q v_p)}, \quad p \in [0, N_a-1].$$

We can represent (1) in vectorized form as

$$\begin{aligned} \tilde{\mathbf{f}} &= \mathbf{T}\mathbf{f}, \quad \mathbf{f} = \mathbf{T}^H \tilde{\mathbf{f}} \\ \text{where } \mathbf{f} &\triangleq [f(l_0, m_0), \dots, f(l_{N-1}, m_{N-1})]^T, \\ \tilde{\mathbf{f}} &\triangleq [\tilde{f}(u_0, v_0), \dots, \tilde{f}(u_{N_a-1}, v_{N_a-1})]^T. \end{aligned} \quad (2)$$

The matrix  $\mathbf{T}$  (size  $N_a \times N$ ) has on its  $q$ -th row and  $p$ -th column  $e^{-j2\pi(l_p u_q + m_p v_q)}$ . Note that given  $\tilde{\mathbf{f}}$ , an estimate for the true image can be formed as  $\hat{\mathbf{f}} = \mathbf{T}^H \tilde{\mathbf{f}}$  but we recover the true image only if  $\mathbf{T}^H \mathbf{T} = \mathbf{I}$ . In order to satisfy this condition we need a regular grid of sampling points in the visibility plane. This is not the case in reality for radio interferometry because the sampling points are determined by location of individual receivers and earth rotation.

## II-B. Prolate spheroidal basis

Our objective is to find a basis function  $p(l, m)$  (and its counterpart in the visibility plane  $\tilde{p}(u, v)$ ) that maximizes the energy in the ROI. The criterion for selection can be written as

$$\lambda = \frac{\sum_{(l,m) \in \text{ROI}} |p(l, m)|^2}{\sum_{(l,m)} |p(l, m)|^2}. \quad (3)$$

In other words,  $\lambda$  in (3) is the ratio between the energy concentrated in the ROI and the total energy in the image. The higher the value of  $\lambda$ , we have lower sidelobes and artifacts outside the ROI.

Let the vectorized versions of  $p(l, m)$  and  $\tilde{p}(u, v)$ , evaluated at the pixels and visibility points, be  $\mathbf{p}$  and  $\tilde{\mathbf{p}}$ , respectively. Then, we can rewrite (3) as

$$\lambda = \frac{\|\mathbf{I}_b^T \mathbf{p}\|^2}{\|\mathbf{p}\|^2} \quad (4)$$

The selection of pixels that belong to the ROI from the full image vector  $\mathbf{p}$  is done by premultiplying by  $\mathbf{I}_b^T$ . Thus,  $\mathbf{I}_b^T$  is an  $N_b \times N$  matrix, constructed by removing rows (corresponding to pixels outside the ROI) from an  $N \times N$  identity matrix.

We state the main result here, which is a direct extension of [7]. The steps required for the removal of an extended source from a given observation are:

- 1) Construct the kernel  $\mathbf{K}$  (size  $N_b$  by  $N_b$ ) as

$$\mathbf{K} = \mathbf{I}_b^T \mathbf{T}^H (\mathbf{T} \mathbf{T}^H)^\dagger \mathbf{T} \mathbf{I}_b \quad (5)$$

- 2) Find the eigendecomposition of  $\mathbf{K}$  and select the eigenmodes with largest eigenvalues. If the  $i$ -th eigenvalue, eigenvector pair of  $\mathbf{K}$  is  $(\lambda_i, \boldsymbol{\eta}_i)$ , the  $i$ -th basis vector is

$$\mathbf{p}_i = \frac{1}{\lambda_i} \mathbf{T}^H (\mathbf{T} \mathbf{T}^H)^\dagger \mathbf{T} \mathbf{I}_b \boldsymbol{\eta}_i \quad (6)$$

- 3) Represent the image as the vector  $\mathbf{b}$  (size  $N$  by 1). Decompose image  $\mathbf{b}$ , using  $M$  basis vectors  $\mathbf{P} \triangleq [\mathbf{p}_0, \dots, \mathbf{p}_{M-1}]$

$$\mathbf{P} \mathbf{m} = \mathbf{b}, \quad \hat{\mathbf{m}} = \mathbf{P}^\dagger \mathbf{b} \quad (7)$$

to find the mode vector  $\hat{\mathbf{m}}$  (size  $M$  by 1).

- 4) Find equivalent basis  $\tilde{\mathbf{P}} = \mathbf{T}^H \mathbf{P}$  in the Fourier plane and subtract the model from the observed data  $\mathbf{z}$ ,

$$\mathbf{r} = \mathbf{z} - \tilde{\mathbf{P}} \hat{\mathbf{m}} \quad (8)$$

to get the residual data vector  $\mathbf{r}$ .

The complete proof of the derivation of PSWF is given in [7]. For completeness, we give a sketch of proof as follows. The numerator of (4) can be written as

$$\begin{aligned} \|\mathbf{I}_b^T \mathbf{p}\|^2 &\stackrel{(a)}{=} \|\mathbf{I}_b^T \mathbf{T}^H \tilde{\mathbf{p}}\|^2 = \text{trace}(\tilde{\mathbf{p}}^H \mathbf{T} \mathbf{I}_b \mathbf{I}_b^T \mathbf{T}^H \tilde{\mathbf{p}}) \\ &\stackrel{(b)}{=} \text{trace}(\boldsymbol{\zeta}^H \mathbf{R}^{-H} \mathbf{T} \mathbf{I}_b \mathbf{I}_b^T \mathbf{T}^H \mathbf{R}^{-1} \boldsymbol{\zeta}) \end{aligned} \quad (9)$$

Here, (a) is obtained by the substitution  $\mathbf{p} = \mathbf{T}^H \tilde{\mathbf{p}}$  and (b) is obtained by using  $\boldsymbol{\zeta} \triangleq \mathbf{R} \tilde{\mathbf{p}}$ , where  $\mathbf{R} \triangleq (\mathbf{T} \mathbf{T}^H)^{1/2}$ . It is easy to simplify the denominator of (4) by substituting  $\mathbf{I}_b = \mathbf{I}$  in (9) as

$$\|\mathbf{p}\|^2 = \text{trace}(\boldsymbol{\zeta}^H \mathbf{R}^{-H} \mathbf{T} \mathbf{T}^H \mathbf{R}^{-1} \boldsymbol{\zeta}) \stackrel{(c)}{=} \text{trace}(\boldsymbol{\zeta}^H \boldsymbol{\zeta}) \quad (10)$$

We used the fact that  $\mathbf{R}^{-H} \mathbf{T} \mathbf{T}^H \mathbf{R}^{-1} = \mathbf{I}$  to obtain (c) in (10).

Using (9) and (10), we can rewrite (4) as

$$\boldsymbol{\zeta} = \arg \max_{\boldsymbol{\zeta}, \|\boldsymbol{\zeta}\|=1} \text{trace}(\boldsymbol{\zeta}^H \mathbf{R}^{-H} \mathbf{T} \mathbf{I}_b \mathbf{I}_b^T \mathbf{T}^H \mathbf{R}^{-1} \boldsymbol{\zeta}) \quad (11)$$

The solution to (11) is the largest eigenvalue, eigenvector pair of the matrix  $\tilde{\mathbf{K}} = \mathbf{R}^{-H} \mathbf{T} \mathbf{I}_b \mathbf{I}_b^T \mathbf{T}^H \mathbf{R}^{-1}$ . The dimension of  $\tilde{\mathbf{K}}$  ( $N_a$  by  $N_a$ ) makes the computation of eigendecomposition prohibitively expensive. In order to reduce this, we apply the transform  $\boldsymbol{\eta} \triangleq \mathbf{I}_b^T \mathbf{T}^H \mathbf{R}^{-1} \boldsymbol{\zeta}$  to  $\tilde{\mathbf{K}} \boldsymbol{\zeta} = \lambda \boldsymbol{\zeta}$  to get

$$\mathbf{I}_b^H \mathbf{T}^H \mathbf{R}^{-2} \mathbf{T} \mathbf{I}_b \boldsymbol{\eta} = \lambda \boldsymbol{\eta} \quad (12)$$

which gives us the kernel  $\mathbf{K}$  in (5) which is of dimension  $N_b$  by  $N_b$ .

## II-C. Computational cost reduction

Under the assumption  $N_a \gg N > N_b$ , the cost of computing (5) involves finding the pseudoinverse  $(\mathbf{T} \mathbf{T}^H)^\dagger$  (size  $N_a$  by  $N_a$ ) whose rank is at most  $N$ . However, the rank of  $\mathbf{K}$  will be at most  $N_b$ . Instead of using  $\mathbf{T}$ , we downsample the visibility points to construct a matrix  $\hat{\mathbf{T}}$  of size  $N_d$  by  $N$  ( $N_d \geq N$ ). This downsampling can be combined with the imaging weights used. Therefore, (5) and (6) are evaluated using  $\hat{\mathbf{T}}$  instead of  $\mathbf{T}$ . However, when we calculate the residual in (8), we use the full matrix  $\mathbf{T}$ .

## III. EXAMPLE

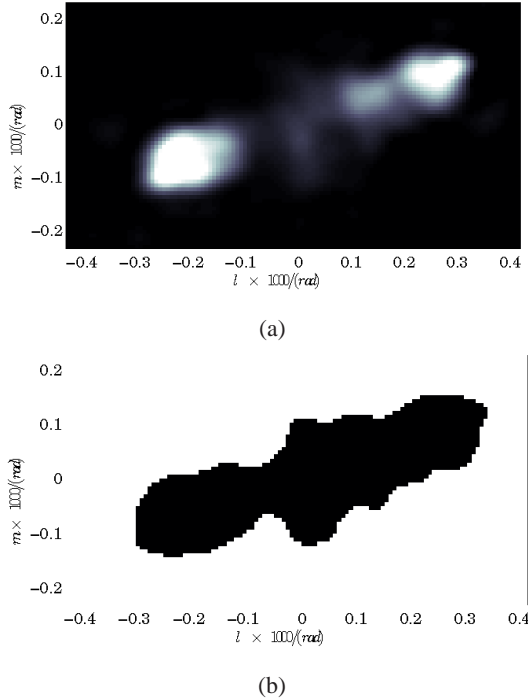
We take a LOFAR (<http://www.lofar.org>) test observation of Cygnus A, at a frequency of 213 MHz as an example. The image of the source Cygnus A is given in Fig. 2(a). The peak flux (not normalized) is about 20 Jy and the total is about 10 kJy. The objective is to subtract this source from the observed data to see the faint background sources. The ROI of this source, which is above the noise floor is given in Fig. 2 (b). The image has  $N = 7552$  pixels of dimension 118 by 64 and the ROI has  $N_b = 1826$  pixels.

The observation lasted for about 8 hours and the original sampling points are shown in Fig. 3 (a). The number of sampling points is  $N_a \approx 15 \times 10^6$ . We downsample the sampling coverage (select only a subset with uniform probability) to get  $N_d = 8931$  which is just above  $N$ . A better approach for downsampling would be to combine this with imaging weights (i.e., selecting fewer short baselines and more long baselines).

The dominant eigenvalues of the kernel  $\mathbf{K}$  (5) (dimension 1826 by 1826) is shown in Fig. 4. We see that approximately the first 100 eigenvalues are close to 1 while any eigenvalue beyond the 200-th eigenmode is almost zero.

In Fig. 5, we have shown the PSWF basis vectors  $\mathbf{p}$  corresponding to the first few largest eigenvalues. Note that the support of the basis vectors are entirely within the ROI, thus creating minimal artifacts outside the ROI. We select the first 100 basis functions  $M = 100$  to construct the matrix  $\mathbf{P}$  in (7). Using this, we decompose the image in Fig. 2 (a) to get the mode vector  $\mathbf{m}$ .

Using the equivalent basis in the Fourier plane, to calculate the residual as in (8). Once the residual is obtained we make the residual images as shown in Fig. 6. For comparison we have also shown the residual image obtained using a shapelet basis function based deconvolution. Both methods give a residual noise level of about 11 mJy far away from Cygnus A. In contrast, with traditional CLEAN based deconvolution, we get a residual noise of about 13 mJy. The shapelet based model used about 300 basis functions of many scales. On the other hand, the PSWF model used only about 100 modes, which is much less.



**Fig. 2.** Cygnus A (a) Observed image at 213 MHz, deconvolved using CLEAN. The peak flux is about 20 Jy and total flux is about 10 kJy. (b) The shaded area correspond to the ROI of 1826 pixels.

#### IV. CONCLUSIONS

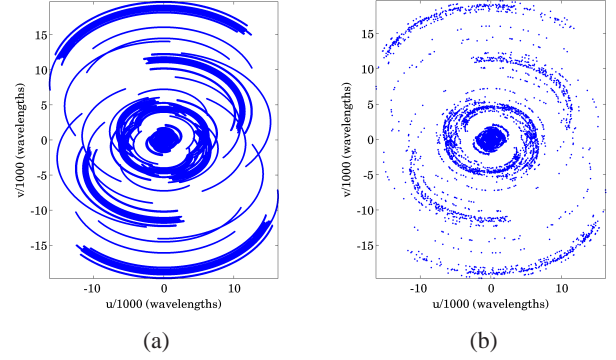
We have presented the use of prolate spheroidal wave functions in radio interferometric image deconvolution and have demonstrated its feasibility by application to a real observation. We get results comparable with existing techniques, but with fewer basis functions and with less artifacts outside the ROI. Future work will focus on widefield imaging and reducing the computational cost.

#### V. ACKNOWLEDGMENTS

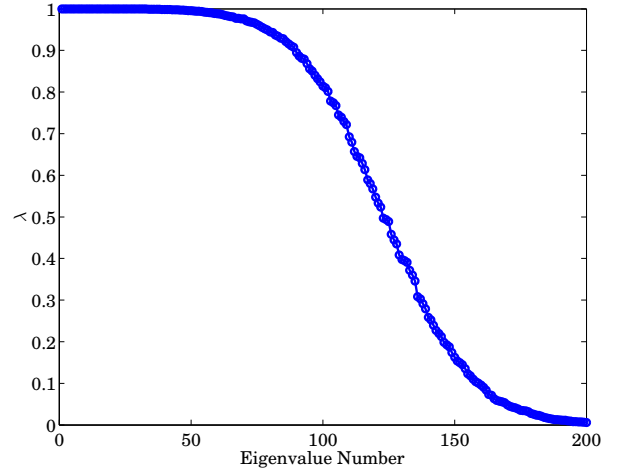
We thank Wim Brouw and Ger de Bruyn for valuable comments and advice.

#### VI. REFERENCES

- [1] S. Yatawatta, "Fundamental limitations of pixel based image deconvolution in radio astronomy," in *proc. IEEE Sensor Array*



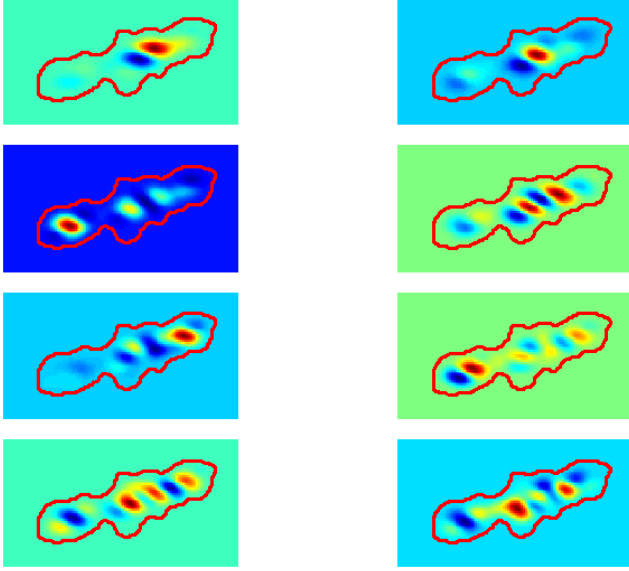
**Fig. 3.** The sampling points in the Fourier plane (a) Original coverage, for an 8 hour observation with about 15 million sampling points. (b) Reduced coverage by downsampling by 1500 with 8931 sampling points. The downsampling is done with uniform probability but it is also possible to combine this with imaging weights, for instance by selecting fewer short baselines and more long baselines.



**Fig. 4.** Eigenspectrum of the kernel  $\mathbf{K}$  with the first 200 largest eigenvalues plotted.

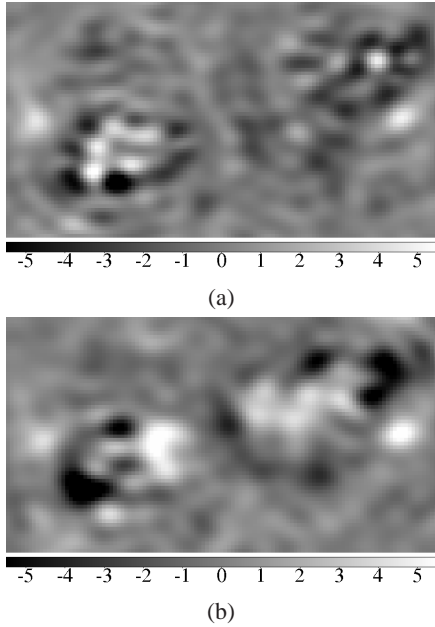
and Multichannel Signal Processing Workshop (SAM), Israel, pp. 69–72, 2010.

- [2] J. Hogbom, "Aperture synthesis with a non regular distribution of interferometer baselines," *A&A Suppl.*, vol. 15, pp. 417–426, 1974.
- [3] R. Levanda and A. Leshem, "Adaptive selective sidelobe canceller beamformer with applications in radio astronomy," in *proc. IEEE 26-th Convention of Electrical and Electronics Engineers (IEEEI)*, Israel, 2010.
- [4] D. Slepian, "On bandwidth," *Proc. of the IEEE*, vol. 64, no. 3, pp. 292–300, 1976.
- [5] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty-I," *Bell Syst. Tech. J.*, vol. 40, no. 2, pp. 43–61, 1961.
- [6] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty-II," *Bell Syst. Tech. J.*, vol. 40, no. 2, pp. 65–84, 1961.
- [7] M. A. Lindquist, C. Zhang, G. Glover, L. Shepp, and



**Fig. 5.** Some PSWF basis vectors corresponding to the largest eigenvalues. The ROI is indicated by the dark curve.

- Q. X. Yang, "A generalization of the two dimensional prolate spheroidal wave function method for nonrectilinear MRI data acquisition methods," *IEEE Trans. on Image Proc.*, vol. 15, no. 9, pp. 2792–2804, 2006.
- [8] W. N. Brouw, "Aperture synthesis," in *Methods in Computational Physics*, vol. 14, pp. 131–175, 1975.
- [9] F. J. Simons and D. V. Wang, "Spatiospectral concentration in the Cartesian plane," *Appl. Comput. Harmon. Anal.* under review, 2010.



**Fig. 6.** Residuals after subtracting Cygnus A using (a) prolate basis (b) shapelet basis. Note that in (b) there is more unsubtracted flux towards the center of the source while in (a) the residuals are compact (point like). Both images have peak values of 5.5 Jy.

